Certificate of Transmission

I hereby certify that this correspondence consisting of 41 pages is being facsimile transmitted to the Patent and Trademark Office on October 7, 2005, to facsimile number (571) 273-8300.

Brian E. Ledell
Reg. No. 42,784

Patent
Attorney's Docket No. 0026-0001

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent Application of	)	**Mail Stop: APPEAL BRIEF - PATENTS**
	)	
Krishna BHARAT et al.	)	Group Art Unit: 2165
	)	
Application No.: 09/729,240	)	Examiner: Hassan Mahmoudi
	)	
Filed: December 5, 2000	)	
	)	
For: IDENTIFICATION OF SEMANTIC UNITS FROM WITHIN A SEARCH QUERY	)	

U.S. Patent and Trademark Office
Customer Window, Mail Stop Appeal Brief Patents
Randolph Building
401 Dulany Street
Alexandria, Virginia 22314

REQUEST FOR REINSTATEMENT OF APPEAL

Sir:

In response to the non-final Office Action, dated April 7, 2005, that reopened prosecution in this application, Appellant respectfully requests reinstatement of the appeal. A Supplemental Appeal Brief accompanies this request.

PATENT
Application No. 09/729,240
Docket No. **0026-0001**

To the extent necessary, a petition for an extension of time under 37 C.F.R. § 1.136 is hereby made. Please charge any shortage in fees due in connection with the filing of this paper, including extension of time fees, to Deposit Account No. 50-1070 and please credit any excess fees to such deposit account.

Respectfully submitted,

HARRITY & SNYDER, L.L.P.

By: _____
Brian E. Ledell
Registration No. 42,784

Date: October 7, 2005

11240 Waples Mill Road
Suite 300
Fairfax, Virginia 22030
(571) 432-0800

RECEIVED
CENTRAL FAX CENTER

OCT 0 7 2005

PATENT
Docket No. **0026-0001**

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE
BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES

In re Patent Application of	**Mail Stop: APPEAL BRIEF - PATENTS**
Krishna Bharat et al.	Group Art Unit: 2165
Application No.: 09/729,240	Examiner: H. Mahmoudi
Filed: December 5, 2000	
For: IDENTIFICATION OF SEMANTIC UNITS FROM WITHIN A SEARCH QUERY	

U.S. Patent and Trademark Office
Customer Window, Mail Stop **Appeal Brief Patents**
Randolph Building
401 Dulany Street
Alexandria, Virginia 22314

SUPPLEMENTAL APPEAL BRIEF

This Supplemental Appeal Brief is submitted in response to the non-final Office Action, dated April 7, 2005, and in support of the Notice of Appeal, filed September 27, 2004.

I. REAL PARTY IN INTEREST

The real party in interest in this appeal is Google Inc.

II. RELATED APPEALS, INTERFERENCES, AND JUDICIAL PROCEEDINGS

Appellants are unaware of any related appeals, interferences or judicial proceedings.

III. STATUS OF CLAIMS

Claims 1-41 are pending in this application. All the claims were rejected in the Office Action of April 7, 2005.

Claims 1-41 stand rejected under U.S.C. § 103(a) as being unpatentable over U.S. Patent No. 6,216,123 to Robertson et al. ("Robertson") in view of U.S. Patent No. 6,295,559 to Emens et al. ("Emens").

Claims 1-41 are the subject of the present appeal. These claims are reproduced in the Claim Appendix of this Appeal Brief.

IV. STATUS OF AMENDMENTS

No amendments have been filed subsequent to the last Office Action, dated April 7, 2005.

Appellants conducted an interview with the Examiner on July 26, 2005. In the interview, Appellants discussed Robertson and Emens and the relevance of these patents to the claims. Independent claim 1 was particularly discussed. Appellants explained that neither Robertson nor Emens were not particularly related to the invention recited in claim 1 and that the rejection of claim 1 under 35 U.S.C. § 103(a) was improper. The Examiner did not agree with the Appellants.

Additionally, in the interview, dependent claims 4, 9, 16, 23, 28, and 33 were discussed. The Examiner indicated that the features of these claims were not disclosed or suggested by the prior art of record.

V. SUMMARY OF CLAIMED SUBJECT MATTER

In the paragraphs that follow, each of the independent claims and the claims reciting means-plus-function or step-plus-function language that is involved in this appeal will be recited followed in parenthesis by examples of where support can be found in the specification and drawings.

Claim 1 is directed to a method of identifying semantic units within a search query. The method includes identifying documents relating to the query (act 202; p. 9, second full paragraph) by comparing search terms in the query to an index of a corpus and generating a plurality of multiword substrings from the query in which each of the substrings includes at least two words (p. 10, lines 6-21). The method further includes calculating, for each of the generated substrings, a value that corresponds to a comparison between one or more of the identified documents and the generated substring (acts 205-207 and acts 301-304; p. 10, lines 14-16; p. 12, lines 8-18). Further, the method includes selecting semantic units from the generated multiword substrings based on the calculated values (acts 208 and 209; p. 10, line 22 through p. 11, line 7).

Claim 6 is directed to a method of locating documents in response to a search query. The method includes receiving the search query from a user (act 201) and generating a list of relevant documents based on search terms of the query (act 202; p. 9, second full paragraph). The method further includes identifying a subset of documents that are most relevant ones of the documents in the list of relevant documents (act 203; p. 10, lines 1-5) and generating a plurality of multiword substrings of the query in which each of the multiword substrings includes at least two words (p. 10, lines 6-21). Still further, the method includes calculating, for each of the generated substrings, a value related to one or more documents in the subset of documents that contain the

substring (acts 205-207 and acts 301-304; p. 10, lines 14-16; p. 12, lines 8-18) and selecting semantic units from the generated multiword substrings based on the calculated value (acts 208 and 209; p. 10, line 22 through p. 11, line 7). Additionally, the method includes refining the generated list of relevant documents based on the selected semantic units (p. 13, lines 10-13).

Claim 11 is directed to a system that includes a server (110) connected to a network (101), the server receives search queries from users via the network. The server includes at least one processor (111) and a memory (112) operatively coupled to the processor. The memory stores program instructions that when executed by the processor, cause the processor to: identify a list of documents relating to the search query by matching individual search terms in the query to an index of a corpus (act 202; p. 9, second full paragraph) generate a plurality of multiword substrings from the query in which each of the substrings includes at least two words (p. 10; lines 6-21); calculate, for each of the generated substrings, a value relating to one or more documents of the identified list of documents that contain the generated substring (acts 205-207 and acts 301-304; p. 10, lines 14-16; p. 12, lines 8-18) and select semantic units from the generated multiword substrings based on the calculated values (acts 208 and 209; p. 10, line 22 through p. 11, line 7).

Claim 18 is directed to a server (110) that includes a processor (111) and a memory (112) operatively coupled to the processor. The memory includes a ranking component (122) configured to return a list of documents ordered by relevance in response to a search query (act 202; p. 9 second full paragraph) and a semantic unit locator component (121) configured to locate semantic units, each having a plurality of words, in search queries entered by a user based on a predetermined number of most relevant documents in the list of documents returned by the

ranking component (acts 204-209 and 301-304; pages 9-13).

Claim 25 is directed to a computer-readable medium (112) storing instructions for causing at least one processor (111) to perform a method that identifies semantic units within a search query. The method includes identifying documents relating to the query by matching individual search terms in the query to an index of a corpus (act 202; p. 9, second full paragraph) and forming a plurality of multiword substrings of the query in which each of the substrings includes at least two words (p. 10, lines 6-21). The method further includes calculating, for each of the substrings, a value relating to the portion of the identified documents that contain the substring (acts 205-207 and acts 301-304; p. 10, lines 14-16; p. 12, lines 8-18). Additionally, the method includes selecting semantic units from the generated multiword substrings based on the calculated values (acts 208 and 209; p. 10, line 22 through p. 11, line 7).

Claim 30 is directed to a computer-readable medium (112) storing instructions for causing a processor (111) to perform a method. The method includes receiving a search query from a user (act 201) and generating a list of relevant documents based on individual search terms of the query (act 202; p. 9, second full paragraph). The method further includes identifying a subset of documents that are the most relevant documents from the list of relevant documents (act 203; p. 10, lines 1-5) and forming a plurality of multiword substrings of the query in which each of the multiword substrings includes at least two words (p. 10, lines 6-21). Additionally, the method includes calculating, for each of the substrings, a value related to the portion of the subset of documents that contain the substring (acts 205-207 and acts 301-304; p. 10, lines 14-16; p. 12, lines 8-18) and selecting semantic units from the generated multiword substrings based on the calculated values (acts 208 and 209; p. 10, line 22 through p. 11, line 7). Further, the

method includes refining the generated list of relevant documents based on the selected semantic units (p. 13, lines 10-13).

Claim 36 is directed to an apparatus (201) for locating documents in response to a search query (act 201). The apparatus comprises means for receiving the search query from a user (110 and act 201) and means for generating a list of relevant documents based on individual search terms of the query (122 and act 202; p. 9, second full paragraph). Further, the apparatus comprises means for identifying a subset of documents that are the most relevant documents from the list of relevant documents (122 and act 203; p. 10, lines 1-5), means for forming a plurality of multiword substrings of the query in which each of the multiword substrings includes at least two words (121 and p. 10, lines 6-21), and means for calculating, for each of the substrings, a value related to the portion of the subset of documents that contain the substring (121 and acts 205-207 and acts 301-304; p. 10, lines 14-16; p. 12, lines 8-18). Further, the apparatus includes means for selecting semantic units from the generated multiword substrings based on the calculated values (121 and acts 208 and 209; p. 10, line 22 through p. 11, line 7) and means for refining the generated list of relevant documents based on the selected semantic units (121 and p. 13, lines 10-13).

Dependent claim 4 depends from claim 3 and further recites discarding generated substrings that overlap other ones of the generated substrings with higher calculated values (Fig. 2, act 209 and page 11, lines 3-5).

Dependent claim 5 depends from claim 1 and further recites that the calculated values are weighted based on a ranking defined by relevance of the identified documents, such that substrings that occur in more relevant ones of the identified documents are assigned higher

calculated values than substrings that occur in less relevant ones of the documents (Fig. 3, acts 301-304 and p. 12, lines 8-15).

Dependent claim 12 depends from claim 11 and recites that the processor refines the identified list of documents based on the selected semantic units (p. 13, lines 10-13).

Dependent claim 21 depends from claim 18 and further recites that the semantic unit locator is further configured to generate a plurality of substrings of the query (p. 10, lines 6-21) and calculate, for each generated substring, a value relating to the portion of the predetermined number of the most relevant documents that contain the substring (acts 205-207 and acts 301-304; p. 10, lines 14-16; p. 12, lines 8-18). Further, as recited in claim 21, the semantic unit locator is configured to locate the semantic units from the generated values (acts 208 and 209; p. 10, line 22 through p. 11, line 7).

Dependent claim 37 depends from claim 1 and further recites that the calculated values are weighted based on a ranking defined by relevance of the identified documents, such that an occurrence of a substring in a more relevant one of the identified documents is weighted more than an occurrence of the substring in a less relevant one of the documents (Fig. 3, acts 301-304 and p. 12, lines 8-15).

VI. GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

A. Claims 1-41 stand rejected under U.S.C. § 103(a) as being obvious over Robertson in view of Emens.

VII. ARGUMENTS

A. The Rejection of Claims 1-41 Under 35 U.S.C. § 103(a) over Robertson in view of Emens Should Be Reversed

1. Definition of Semantic Unit As Used In Pending Claims

Claim 1 is directed to a method of identifying *semantic units*. The method includes a number of acts that, among other things, generate a plurality of multiword substrings and then select semantic units from the generated multiword substrings based on the calculated values. Each of the other independent claims also recites the phrase *semantic unit* or *semantic units*.

In previous Office Actions and responses to Office Actions, Appellants and the Examiner have disagreed over the meaning of the phrase "semantic unit." Appellants submit that the term "semantic unit," as defined by the Appellants' specification, refers to multiple terms that are considered to function as a "compound" that forms a single, semantically meaningful unit. (See Spec., page 2). In previous Office Actions, the Examiner has refused to use this definition, stating "multiple terms that are considered to function as a 'compound' that forms a single semantically meaningful unit is not recited in the rejected claim." (Final Office Action of May 28, 2004, page 14). In previous responses, the Examiner appeared to be interpreting "semantic units" very broadly to cover virtually any text string(s). (See final Office Action of May 28, 2004, pages 2 and 14).

Appellants submit that interpreting the phrase "semantic units" to cover virtually any text string is overly broad and is inapposite to the plain meaning of the phrase. The Merriam-Webster Online dictionary, for instance, defines semantic as "of or relating to meaning in language." Thus, a multiword semantic unit, as recited in claim 1, refers to multiple terms related by meaning.

Additionally, Appellants note that an applicant is entitled to be his or her own lexicographer. *See In re Paulsen*, 30 F.3d 1475, 1480 31 USPQ2d 1671, 1674 (Fed. Cir. 1994). In this regard, Appellants' specification clearly defines and uses the term semantic unit consistent with the plain meaning of the phrase. At page 2, for instance, Appellants' specification defines the term semantic unit (also called a compound in the specification) in the context of the example semantic unit "baldur's gate":

> Multiple search terms entered by a user are often more useful if considered by the search engine as a single compound unit. Assume that a user enters the search terms "baldur's gate download." The user intends for this query to return web pages that are relevant to the user's intention of downloading the computer game called "baldur's gate." Although "baldur's gate" includes two words, the two words together form a single semantically meaningful unit. If the search engine is able to recognize "baldur's gate" as a single semantic unit, called a compound herein, the search engine is more likely to return the web pages desired by the user.

Page 4 of Appellants' specification further elaborates on this definition:

> For example, the queries "country western mp3" and "leaving the old country western migration" both have the words "country" and "western" next to each other. Only for the first query, however, is "country western" a representative compound. Segmenting such queries correctly requires some understanding of the meaning of the query. In the second query, the compound "western migration" is more appropriate, although it occurs less frequently in general.

In summary, Appellants submit that under a reasonable interpretation of the phrase "semantic unit," a semantic unit refers to two or more terms that function as a "compound" that forms a single, semantically meaningful unit.

2. Rejection of Claims 1, 2, 11, 14, 25, 26

As described in Appellants' specification, identifying semantic units within a search

query, such as performed via the elements of claim 1, can be important in an application such as a search engine to, for example, modify how ranked results are returned to the user based on the presence of the semantic units or to suggest alternate queries to the user. (Spec., pages 2 and 3). Previous approaches to detecting semantic units include pre-extracting semantic units from a document corpus and extracting semantic units from a query log. (Spec., see last full paragraph on page 3 and the first full paragraph on page 4). As is further described in the specification, a disadvantage of these previous approaches is that they tend to ignore the meaning of the query in which the semantic unit occurs (i.e., they ignore the context in which the semantic unit occurs). (Spec., page 4). For example, the queries "country western mp3" and "leaving the old country western migration" both have the words "country" and "western" next to each other. Only for the first query, however, is "country western" a semantic unit. (Spec., paragraph bridging pages 4 and 5).

As discussed above, claim 1 is a method of identifying semantic units within a search query. Claim 1 includes identifying documents relating to the query by comparing search terms in the query to an index of a corpus and generating a plurality of multiword substrings from the query in which each of the substrings includes at least two words. Claim 1 further recites calculating, for each of the generated substrings, a value that corresponds to a comparison between one or more of the identified documents and the generated substring and selecting semantic units from the generated multiword substrings based on the calculated values.

It is a cardinal tenant of patent law that to establish a *prima facie* obviousness of a claimed invention, all the claim limitations must be taught or suggested by the prior art. *In re Royka*, 490 F.2d 981, 180 USPQ 580 (CCPA 1974). "All words in a claim must be considered in

judging the patentability of that claim against the prior art." *In re Wilson*, 424 F.2d 1382, 1385, 165 USPQ 494, 496 (CCPA 1970). If an independent claim is nonobvious under 35 U.S.C. § 103, then any claim depending therefrom is nonobvious. *In re Fine*, 837 F.2d 1071, 5 USPQ2d 1596 (Fed. Cir. 1988).

In rejecting representative claim 1, the Examiner contends that Robertson discloses the first three elements recited in claim 1, but concedes that Robertson does not disclose the last element of claim 1. (Office Action of April 7, 2005, page 3). The Examiner contends, however, that Emens cures the deficiencies of Robertson and states that it would have been obvious to modify Robertson in view of Emens to disclose the invention recited in claim 1. Appellants strongly disagree with the Examiner's assertions. In particular, as will be discussed below, Robertson fails to disclose or suggest many of the elements recited in claim 1. Emens is similarly deficient and does not disclose or suggest the element of claim 1 that the Examiner concedes is not disclosed by Robertson. Thus, all of the claimed limitations are not taught or suggested and the rejection of this claim should be reversed.

Robertson is directed to techniques for generating and searching a full text index for a search engine. (Robertson, Abstract). The index of Robertson is said to be "extremely efficient and greatly reduces table accesses and/or disk I/Os." (Robertson, Abstract). More particularly, Robertson discloses associating words with "word numbers" when indexing documents in a manner that "greatly simplifies and reduces the overhead involved in the determination of whether multiple search words exist in the same document." (Robertson, col. 2, line 64 through col. 3, line 3).

Although Robertson discloses a search engine and the matching of a search query to a

document corpus, Robertson does not disclose or suggest, as is recited in claim 1, "generating a plurality of multiword substrings from the query in which each of the substrings includes at least two words." (emphasis added). The Examiner contends that Robertson discloses this feature and points to a number of sections of Robertson, including column 4, line 63 through column 5, line 5; column 8, lines 12-23; column 2, lines 52-56; and column 13, lines 15-21. (Office Action of April 7, 2005, page 3). Applicants respectfully disagree with the Examiner's interpretation of Robertson.

Column 4, line 63 through column 5, line 5 of Robertson discloses:

> Each word number cluster includes one or more word numbers that have been combined during a word register operation, and which therefor satisfy a search operation, such as a proximity operation, and can thereafter be treated as a single word number. A word cluster has a single relevance number associated with it, eliminating the need to repeatedly process multiple independent word relevance numbers. Treating clusters of word numbers as one unit allows the attachment of a single relevance value to a semantic unit.

This section of Robertson relates to forming a "word number cluster" during the process of obtaining results for a search operation. As previously mentioned, Robertson associates word numbers with words and documents. The word numbers that are part of the word number clusters of Robertson relate to words/documents identified as part of the process of obtaining results for documents. These word number clusters of Robertson, however, cannot be said to correspond to, as recited in claim 1, generating a plurality of multiword substrings from the query.

In addition to column 4, line 63 through column 5, line 5, the Examiner pointed to a number of other sections of Robertson as allegedly disclosing "generating a plurality of multiword substrings from the query in which each of the substrings includes at least two

words," as recited in claim 1. These other sections of Robertson also fail to disclose or suggest this feature of claim 1. Column 8, lines 12-23 of Robertson, for instance, generally discusses the association of word numbers to indexed documents. Column 2, lines 52-56 relate a group of word numbers, which Robertson associates with each document being indexed. Column 13, lines 15-21 relates to cross-referencing documents and groups of numbers. All of these sections of Robertson describe some aspect of the document indexing scheme used by Robertson. None of these sections, however, nor any other portion of Robertson, disclose or suggest generating a plurality of multiword substrings from the query in which each of the substrings includes at least two words.

Claim 1 additionally recites calculating, for each of the generated substrings, a value that corresponds to a comparison between one or more of the identified documents and the generated substring. At least because Robertson does not disclose or suggest generating a plurality of multiword substrings, Appellants submit that Robertson could not possibly disclose or suggest calculating a value for each of the generated substrings, as is also recited in claim 1. The Examiner contends that Robertson discloses this feature of claim 1 and cites column 14, lines 9-64 and column 16, line 50 through column 17, line 8. (Office Action of April 7, 2005, page 3). These sections of Robertson generally relate to identifying relevant documents based on search queries. Appellants do not dispute that Robertson discloses a search engine that identifies and ranks relevant documents based on search queries. The search engine of Robertson, however, appears to operate on a conventional search query received from a user and is applied to a standard document corpus. In contrast, the values calculated by this feature of claim 1 are calculated for each of the generated substrings (as generated in the second element of claim 1)

based on a comparison between one or more of the identified documents (as identified in the first element of claim 1).

Claim 1 additionally recites selecting semantic units from the generated multiword substrings based on the calculated values. The Examiner relies on Emens to allegedly disclose this feature of claim 1. (Office Action of April 7, 2005, page 4). Appellants respectfully disagree with the Examiner's interpretation of Emens.

Emens is directed to rating hypermedia content by rating the content based on a degree of objectionable content. (Emens, Title and Abstract). Emens recognizes the concept of a semantic unit. Emens, for instance, states: "Raw data file 70 is parsed in step 72 into semantic units 74, which may be words, phrases, or other text groupings. Parsing text data into words or phrases is a well-known technique." (Emens, column 5, lines 31-34).

Although Emens discloses parsing a data file into semantic units, Emens does not disclose any particular technique for parsing the semantic units. In fact, Emens explicitly states that "parsing text data into words or phrases is a well-known technique." Conventional techniques for locating semantic units are also described in the Background of the Invention section of Appellants' specification, at, for example, pages 3 and 4.

Because Emens does not disclose any particular technique for identifying semantic units, Emens could not possible disclose or suggest selecting semantic units from the generated multiword substrings based on the calculated values, as recited in claim 1. The Examiner points to column 5, lines 28-48 and column 6, line 54 through column 7, line 5 of Emens as allegedly disclosing this aspect of the invention recited in claim 1. A portion of the cited section of column 5 is discussed in the previous paragraph. Although this section of Emens mentions

semantic units, this section of Emens does not disclose any specific technique for identifying semantic units. Column 6, line 54 though column 7, line 5 of Emens discloses, among other things, operations by which a search engine produces a "rated search result page." This section of Emens, however, also fails to disclose any specific technique for identifying semantic units, much less selecting semantic units from the generated multiword substrings based on the calculated values, as recited in claim 1.

For at least the foregoing reasons, Appellants submit that Robertson and Emens, even if combined as the Examiner suggests, do not disclose or suggest many of the features recited in claim 1.

Furthermore, Appellants submit that the Examiner has not made a *prima facie* case of obviousness with regard to claim 1. The initial burden of establishing a *prima facie* basis to deny patentability to a claimed invention always rests upon the Examiner. In re Oetiker, 977 F.2d 1443, 24 USPQ2d 1443 (Fed. Cir. 1992). In rejecting a claim under 35 U.S.C. § 103, the Examiner must provide a factual basis to support the conclusion of obviousness. In re Warner, 379 F.2d 1011, 154 USPQ 173 (CCPA 1967). Based upon the objective evidence of record, the Examiner is required to make the factual inquiries mandated by Graham v. John Deere Co., 86 S.Ct. 684, 383 U.S. 1, 148 USPQ 459 (1966). The Examiner is also required to explain how and why one having ordinary skill in the art would have been realistically motivated to modify an applied reference and/or combine applied references to arrive at the claimed invention. Uniroyal, Inc. v. Rudkin-Wiley Corp., 837 F.2d 1044, 5 USPQ2d 1434 (Fed. Cir. 1988).

As discussed above, neither Robertson nor Emens discloses any particular technique for identifying semantic units. Robertson is primarily directed to an improved indexing system for a

search engine. Emens is primarily directed to rating content for objectionable material. At most, Emens discloses identifying semantic units using well-known techniques. (Emens, column 5, lines 31-34). Appellants submit that because neither Robertson nor Emens disclose or suggest any particular technique for identifying semantic units, it is entirely implausible to argue, as the Examiner is arguing, that one of ordinary skill in the art would somehow, upon reading these references, be motivated to identify semantic units using the specific techniques recited in claim 1. Accordingly, the Examiner has not made a *prima facie* case of obviousness with regard to claim 1.

For at least these reasons, Appellants submit that the rejection of claim 1 is improper and should be reversed.

3. Claims 4, 9, 16, 23, 28, and 33

Claim 4 depends from claim 3, which depends from claim 1. In the interview conducted with Appellants' representative on August 1, 2005, the Examiner verbally indicated that claims 4, 9, 16, 23, 28, and 33 were directed to allowable subject matter. Because this indication has not been made of record in a formal Office Action, however, Appellant will assume, for the purpose of this Appeal, that these claims are still officially rejected.

Claim 4 recites that the selection of the semantic units further includes discarding the generated substrings that overlap other ones of the generated substrings with higher calculated values. Neither Robertson nor Emens in any way disclose or suggest these features. The Examiner pointed to column 19, line 19 through column 20, line 20 and column 21, lines 1-56 of Robertson as disclosing the features of claim 4. (Office Action of April 7, 2005, page 5). These

sections of Robertson relate to the description of the flow charts of Figs. 11 and 12 of Robertson. These flowcharts are respectively described by Robertson as a flow diagram illustrating block 106 of Fig. 9 ("Merge Valid Hits From Reg1 and Reg2") and a flow diagram illustrating a process for combining two word registers into a result word register during an 'OR' operation. Nowhere do these flowcharts mention or in any way suggest discarding substrings, much less discarding the generated substrings of claim 4 when the substrings overlap other ones of the generated substrings with higher calculated values, as is also recited in claim 4.

For at least these reasons, Appellants submit that the rejection of claim 4 is improper and should be reversed.

4. Claims 6, 7, 30, 31, 35, and 36

Claim 6 is directed to a method of locating documents in response to a search query. Claim 6 recites a number of features similar to those recited in claim 1, including "generating a plurality of multiword substrings of the query" and "selecting semantic units from the generated multiword substrings based on the calculated values." For reasons similar to those given above regarding claim 1, Appellants submit that Robertson and Emens, either taken alone or in combination, do not disclose or suggest these features of claim 6.

Also, Appellants assert that one of ordinary skill in the art would not be motivated to combine Robertson and Emens to obtain the invention of claim 6. Neither Robertson nor Emens is particularly concerned with identifying semantic units. Accordingly, one of ordinary skill in the art would not be motivated to create a technique for identifying semantic units, much less the specific technique recited in claim 6 for selecting semantic units and using the selected semantic

units to refine search results. The Examiner is merely picking and choosing various isolated phrases that are at most tangentially related to the features of claim 6 and then reconstructing the features of claim with an analysis based purely on hindsight gleaned from Applicants' specification.

Claim 6 includes additional features not disclosed or suggested by Robertson and Emens, either alone or in combination. For instance, claim 6 recites "identifying a subset of documents that are most relevant ones of the documents in the list of relevant documents." The Examiner contends that Robertson, at column 14, lines 9-24, discloses this feature of claim 6. (Office Action of April 7, 2005, page 6). This section of Robertson discloses:

> As will be discussed in greater detail herein, one embodiment of the present invention maintains and calculates relevance information such that documents that match a search request can be ranked in order of importance. In general, two types of relevance are maintained, attribute relevance and processing relevance. Attribute relevance relates primarily to static attributes of words, and is collected during the generation of the full text index. Such static relevance information can comprise, for example, whether a word appears in a title, is bolded, was italicized, or offset in some special manner. Processing relevance relates to relationships of the words to each other, such as the proximity of one word to another word in a search request, or the number of occurrences of the word in a document. Processing relevance values are generated when an operation is being applied against a word register.

(Robertson, column 14, lines 9-24). This section of Robertson appears to generally discuss relevance values for documents. In no way, however, could this section of Robertson be considered to disclose or suggest identifying a subset of documents that are most relevant ones of the documents in the list of relevant documents, as recited in claim 6.

Claim 6 additionally recites "refining the generated list of relevant documents based on the selected semantic units," in which the list of relevant documents is generated based on search

terms of a query. Neither Robertson nor Emens identify semantic units present in a search query, and accordingly, they could not possibly disclose refining a list of documents based on selected semantic units. The Examiner, however, contends that this feature of claim 6 is disclosed by Emens at column 6, line 54 through column 7, line 25. (Office Action of April 7, 2005, , page 7). This section of Emens states:

> The embodiment of the present invention for rating search result pages may be implemented in a distributed computer system in various ways. FIG. 6 is a block diagram showing one potential embodiment. A user of a client browser 130 sends a search query 132 to a search engine 134. Upon receiving search query 132, search engine 134 performs a search of index 137 in step 136 to generate a raw search result page 138. Search engine 134 derives a CCRV 144 for raw search result page 138 and stores it to produce rated search result page 142 in step 140. Search engine 134 sends rated SRP 142 to client browser 130, which uses CCRV 144 to determine whether or not to display SRP 142 to the user. In step 146, client browser 130 compares CCRV 144 with preset user limit values 148. If one component of CCRV 144 is greater than the corresponding preset user limit value 148, client browser 130 does not display SRP 142 (step 150). Alternately (step 152), it does display SRP 142.
>
> In an alternate embodiment shown in FIG. 7, the decision to display the search result page is made by the search engine rather than by the browser. In this case, client browser 160 sends both search query 162 and preset user limit values 164 to search engine 165. As before, search engine 165 performs a search (step 166) of index 168 to create raw search result page 170. It then derives a CCRV 176, which it stores to produce a rated search result page 174 in step 172. In step 178, search engine 165 determines whether or not to send rated SRP 174 to client browser 160 by comparing CCRV 176 with preset user limit values 164. If one component of CCRV 176 is too high (step 180), search engine 165 does not send SRP 174, instead sending an explanation of why it cannot send the page. Alternately (step 182), it does send SRP 174, and client browser 160 displays the page (step 184), because its rating is necessarily below preset user limit values 164.

(Emens, column 6, line 54 through column 7, line 25). This section of Emens completely lacks any disclosure of semantic units or refining a search query, much less, as is recited in claim 6, refining a generated list of relevant documents based on selected semantic units.

For at least the foregoing reasons, Appellants submit that Robertson and Emens do not disclose or suggest each of the features recited in claim 6. Reversal of the rejection of claim 6 is therefore respectfully requested.

5. Claim 18

Independent claim 18 recites a number of features, including a ranking component configured to return a list of documents ordered by relevance in response to a search query and a semantic unit component configured to locate semantic units, having a plurality of words, in search queries entered by a user based on a predetermined number of most relevant documents in the list of documents returned by the ranking component. As previously discussed, neither Robertson nor Emens discloses or suggests locating semantic units in search queries, much less locating semantic units based on a predetermined number of most relevant documents in a list of documents returned by the ranking component. Accordingly, Appellants submit that neither Robertson nor Emens, either alone or in combination, could possibly disclose or suggest the semantic unit locator component recited in claim 18.

For at least the foregoing reasons, Appellants submit that Robertson and Emens do not disclose or suggest each of the features recited in claim 18. Accordingly, the rejection of claim 18 under 35 U.S.C. § 103(a) in view of Robertson and Emens is improper and should be reversed.

6. Claims 3, 8, 15, 22, 27, and 32

Dependent claim 3 recites that the selection of the semantic units further includes

selecting semantic units from the generated substrings that have calculated values above a predetermined threshold. The Examiner points to column 5, lines 22-27 and column 20, lines 44-67 of Robertson as disclosing this feature. (Office Action of April 7, 2005, page 5). These sections of Robertson state:

> According to still another embodiment of the present invention relevance values are "self-normalizing" and are maintained in a predetermined relevance value range. Document level relevance values are independent of any other document level relevance values associated with the documents, and can thus be returned to a user immediately.

(Robertson, column 5, lines 22-27)

> The relevance calculations of the present invention are preferably self-normalizing. In other words, relevance values exist within a predetermined range and do not need to be normalized against other relevance values before they can be returned to a user. For example, relevance values can be maintained in a predetermined range between about 0 and about 100. Cumulative relevance values, that is relevance values created as a function of other relevance values, such as cluster relevance values or document level relevance values all exist in this predetermined range, and thus are independent of one another. Due to this self-normalizing feature, results can be returned `piece-meal` from a search and immediately displayed to a user with a usable relevance value even though the search request is perhaps still awaiting results of the search from certain remote servers. In conventional searching engines, it is typically necessary to normalize the relevance values associated with matching documents. Consequently, results cannot be returned to a user until all of the results have been accumulated, and the relevance values have all been normalized. The present invention eliminates the need to normalize relevance values among documents because each relevance calculation is self-normalizing and maintains a relevance value in a predetermined range from about 0 to about 100.

(Robertson, column 20, lines 44-67). These sections of Robertson relate to self-normalizing relevance values. These sections of Robertson in no way disclose or suggest, however, selecting semantic units from the generated substrings that have calculated values above a predetermined threshold. Indeed, Appellants fail to see how a self-normalizing relevance value for a document

is in anyway related to selecting semantic units from generated substrings that have calculated values above a predetermined threshold, as recited in claim 3.

For at least this reason, in addition to the fact that claim 3 depends from claim 1, Appellants submit that the rejection of claim 3 is improper and should be withdrawn.

7. Claims 5, 10, 17, 24, 29, and 34

Dependent claim 5 further defines the features of claim 1 and recites that "the calculated values are weighted based on a ranking defined by relevance of the identified documents, such that substrings that occur in more relevant ones of the identified documents are assigned higher calculated values than substrings that occur in less relevant ones of the documents."

The Examiner points to column 14, lines 33-45 and column 1, lines 20-25 of Robertson as disclosing these features. (Office Action of April 7, 2005, page 5). These sections of Robertson discuss techniques for generating a combined relevance value based on partial relevance results, such as a technique for combining the relevance of a first word to a document and the relevance of a second word to the document to obtain a final relevance value of both words to the document. Calculating a combined relevance value for a document, however, as disclosed by Robertson, does not disclose or suggest the features recited in claim 5, in which substrings that occur in more relevant ones of the identified documents are assigned higher calculated values than substrings that occur in less relevant ones of the documents. Robertson completely fails to disclose or suggest this feature.

Accordingly, for these reasons, the rejection of claim 5 should also be reversed.

8. Claim 21

Claim 21 depends from claim 18, and further recites that the semantic unit locator is further configured to, *inter alia*, calculate, for each generated substring, a value relating to the portion of the predetermined number of the most relevant documents that contain the substring; and locate the semantic units from the generated values. As discussed above, although Robertson may disclose a search engine that calculates relevance values that each relate a document to a search query, Robertson completely fails to disclose or suggest calculating the value recited in claim 21. Appellants submit that Emens also completely fails to disclose or suggest these features of claim 21.

For at least the foregoing reasons, Appellants submit that Robertson and Emens do not disclose or suggest each of the features recited in claim 21. Accordingly, the rejection of claim 21 under 35 U.S.C. § 103(a) in view of Robertson and Emens is improper and should be reversed.

9. Claims 12, 13, 19, and 20

Claims 12, 13, 19, and 20 are dependent claims. Representative claim 12 recites that a processor refines the identified list of documents based on the selected semantic units. The Examiner cites column 6, line 54 through column 7, line 25 of Emens as allegedly disclosing this feature. (Office Action of April 7, 2005, page 8). This section of Emens was discussed above in regard to independent claim 6. As mentioned, this section of Emens completely lacks any disclosure of semantic units or refining a search query, much less, as is recited in claim 12, refining an identified list of documents based on selected semantic units.

For at least these reasons, Appellants submit that neither Robertson nor Emens, either alone or in combination, disclose or suggest the features recited in claim 12. Accordingly, the rejection of claims 12, 13, 19, and 20 under 35 U.S.C. § 103(a) should be reversed.

10. Claims 37-41

Claims 37-41 are dependent claims. Representative claim 37 recites that the calculated values are weighted based on a ranking defined by relevance of the identified documents, such that an occurrence of a substring in a more relevant one of the identified documents is weighted more than an occurrence of the substring in a less relevant one of the documents. The Examiner alleges that this feature is disclosed by Robertson, and particularly points to column 7, line 57 through column 8, line 10; column 14, lines 9-64; and column 15, lines 45-51. These sections of Robertson generally relate to the operation of a search engine in calculating relevance values of a search query to documents. Although claim 37 does include the words "relevance" and "documents," claim 37 does not simply recite determining document relevance to a search query. More specifically, claim 37 recites that the "calculated values are weighted . . . such that an occurrence of a substring in a more relevant one of the identified documents is weighted more than an occurrence of the substring in a less relevant one of the documents." Robertson completely fails to disclose or suggest any such weighting of the calculated values recited in claim 37.

Accordingly, Appellants submit that neither Robertson nor Emens, either alone or in combination, disclose or suggest the features recited in claim 37. Accordingly, the rejection of claim 37 under 35 U.S.C. § 103(a) should be reversed.

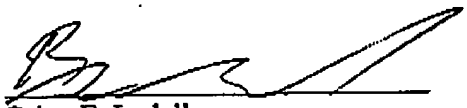
VIII. CONCLUSION

In view of the foregoing arguments, Appellants respectfully solicit the Honorable Board to reverse the Examiner's rejections of claims 1-41 under 35 U.S.C. § 103(a).

To the extent necessary, a petition for an extension of time under 37 C.F.R. § 1.136 is hereby made. Please charge any shortage in fees due in connection with the filing of this paper, including extension of time fees, to Deposit Account No. 50-1070 and please credit any excess fees to such deposit account.

Respectfully submitted,

HARRITY & SNYDER, L.L.P.



Brian E. Ledell
Reg. No. 42,784

Date: October 7, 2005

11240 Waples Mill Road
Suite 300
Fairfax, Virginia 22030
(571) 432-0800

Customer No. 44989

IX. CLAIM APPENDIX

1. A method of identifying semantic units within a search query comprising:

identifying documents relating to the query by comparing search terms in the query to an index of a corpus;

generating a plurality of multiword substrings from the query in which each of the substrings includes at least two words;

calculating, for each of the generated substrings, a value that corresponds to a comparison between one or more of the identified documents and the generated substring; and

selecting semantic units from the generated multiword substrings based on the calculated values.

2. The method of claim 1, wherein the identification of the documents further includes:

generating an initial list of relevant documents; and

selecting a predetermined number of most relevant ones of the documents in the initial list as the identified documents.

3. The method of claim 1, wherein the selection of the semantic units further includes:

selecting semantic units from the generated substrings that have calculated values above a predetermined threshold.

4. The method of claim 3, wherein the selection of the semantic units further includes:

discarding the generated substrings that overlap other ones of the generated substrings with higher calculated values.

5. The method of claim 1, wherein the calculated values are weighted based on a ranking defined by relevance of the identified documents, such that substrings that occur in more relevant ones of the identified documents are assigned higher calculated values than substrings that occur in less relevant ones of the documents.

6. A method of locating documents in response to a search query, the method comprising:

receiving the search query from a user;

generating a list of relevant documents based on search terms of the query;

identifying a subset of documents that are most relevant ones of the documents in the list of relevant documents;

generating a plurality of multiword substrings of the query in which each of the multiword substrings includes at least two words;

calculating, for each of the generated substrings, a value related to one or more documents in the subset of documents that contain the substring;

selecting semantic units from the generated multiword substrings based on the calculated values; and

refining the generated list of relevant documents based on the selected semantic units.

7. The method of claim 6, wherein the identified subset includes a predetermined number of the most relevant ones of the documents in the list of relevant documents.

8. The method of claim 6, wherein the selection of the semantic units further includes:

selecting semantic units from the generated substrings that have calculated values above a predetermined threshold.

9. The method of claim 8, wherein the selection of the semantic units further includes:

discarding the generated substrings that overlap other ones of the generated substrings with higher calculated values.

10. The method of claim 6, wherein the calculated values are weighted based on a ranking defined by relevance of the identified documents, such that substrings that occur in more relevant ones of the documents are assigned higher calculated values than substrings that occur in less relevant ones of the documents.

11. A system comprising:

a server connected to a network, the server receiving search queries from users via the

network, the server including:

at least one processor; and

a memory operatively coupled to the processor, the memory storing program instructions that when executed by the processor, cause the processor to: identify a list of documents relating to the search query by matching individual search terms in the query to an index of a corpus; generate a plurality of multiword substrings from the query in which each of the substrings includes at least two words; calculate, for each of the generated substrings, a value relating to one or more documents of the identified list of documents that contain the generated substring; and select semantic units from the generated multiword substrings based on the calculated values.

12. The system of claim 11, wherein the processor refines the identified list of documents based on the selected semantic units.

13. The system of claim 12, wherein the system transmits the refined list of documents to the user.

14. The system of claim 11, wherein the network is the Internet and the corpus is a collection of web documents.

15. The system of claim 11, wherein the memory includes instructions for causing the processor to:

select semantic units from the generated substrings that have calculated values above a predetermined threshold.

16. The system of claim 15, wherein the memory includes instructions for causing the processor to:

discard substrings that overlap other substrings with a higher calculated value.

17. The system of claim 11, wherein the calculated values are weighted based on a ranking defined by relevance of the identified documents, such that substrings that occur in more relevant documents are assigned higher calculated values than substrings that occur in less relevant documents.

18. A server comprising:

a processor; and

a memory operatively coupled to the processor, the memory including:

a ranking component configured to return a list of documents ordered by relevance in response to a search query; and

a semantic unit locator component configured to locate semantic units, each having a plurality of words, in search queries entered by a user based on a predetermined number of most relevant documents in the list of documents returned by the ranking component.

19. The server of claim 18, further including:

a search engine configured to refine the list of documents based on the located semantic units.

20. The server of claim 19, wherein the processor is configured to:

transmit the refined list of documents to a user that provided the query.

21. The server of claim 18, wherein the semantic unit locator is further configured to:

generate a plurality of substrings of the query;

calculate, for each generated substring, a value relating to the portion of the predetermined number of the most relevant documents that contain the substring; and

locate the semantic units from the generated values.

22. The server of claim 21, wherein the semantic unit locator is configured to locate semantic units from the generated substrings that have calculated values above a predetermined threshold.

23. The server of claim 22, wherein the semantic unit locator is configured to discard substrings that overlap other substrings with a higher calculated value.

24. The server of claim 21, wherein the calculated values are weighted based on a ranking defined by relevance of the identified documents, such that substrings that occur in more relevant documents are assigned higher calculated values than substrings that occur in less

relevant documents.

25. A computer-readable medium storing instructions for causing at least one processor to perform a method that identifies semantic units within a search query, the method comprising:

identifying documents relating to the query by matching individual search terms in the query to an index of a corpus;

forming a plurality of multiword substrings of the query in which each of the substrings includes at least two words;

calculating, for each of the substrings, a value relating to the portion of the identified documents that contain the substring; and

selecting semantic units from the generated multiword substrings based on the calculated values.

26. The computer-readable medium of claim 25, wherein the identification of the set of documents further includes:

generating an initial list of relevant documents; and

selecting a predetermined number of the most relevant documents in the initial list to include in the identified documents.

27. The computer-readable medium of claim 25, wherein the selection of the semantic units further includes:

selecting semantic units from the generated substrings that have calculated values above a predetermined threshold.

28. The computer-readable medium of claim 27, wherein the selection of the semantic units further includes:

discarding substrings that overlap other substrings with a higher calculated value.

29. The computer-readable medium of claim 27, wherein the calculated values are weighted based on a ranking defined by relevance of the identified documents, such that substrings that occur in more relevant documents are assigned higher calculated values than substrings that occur in less relevant documents.

30. A computer-readable medium storing instructions for causing a processor to perform a method, the method comprising:

receiving the search query from a user;

generating a list of relevant documents based on individual search terms of the query;

identifying a subset of documents that are the most relevant documents from the list of relevant documents;

forming a plurality of multiword substrings of the query in which each of the multiword substrings includes at least two words;

calculating, for each of the substrings, a value related to the portion of the subset of documents that contain the substring;

selecting semantic units from the generated multiword substrings based on the calculated values; and

refining the generated list of relevant documents based on the selected semantic units.

31. The computer-readable medium of claim 30, wherein the identified subset includes a predetermined number of the most relevant documents from the list of relevant documents.

32. The computer-readable medium of claim 30, wherein the selection of the semantic units further includes:

selecting semantic units from the generated substrings that have calculated values above a predetermined threshold.

33. The computer-readable medium of claim 32, wherein the selection of the semantic units further includes:

discarding substrings that overlap other substrings with a higher calculated value.

34. The computer-readable medium of claim 30, wherein the calculated values are weighted based on a ranking defined by relevance of the identified documents, such that substrings that occur in more relevant documents are assigned higher calculated values than substrings that occur in less relevant documents.

35. The computer-readable medium of claim 30, wherein the computer-readable medium is a CD-ROM, floppy disk, tape, flash memory, system memory, hard drive, or data signal embodied in a carrier wave.

36. An apparatus for locating documents in response to a search query, comprising:

means for receiving the search query from a user;

means for generating a list of relevant documents based on individual search terms of the query;

means for identifying a subset of documents that are the most relevant documents from the list of relevant documents;

means for forming a plurality of multiword substrings of the query in which each of the multiword substrings includes at least two words;

means for calculating, for each of the substrings, a value related to the portion of the subset of documents that contain the substring;

means for selecting semantic units from the generated multiword substrings based on the calculated values; and

means for refining the generated list of relevant documents based on the selected semantic units.

37. The method of claim 1, wherein the calculated values are weighted based on a ranking defined by relevance of the identified documents, such that an occurrence of a substring in a more relevant one of the identified documents is weighted more than an occurrence of the

substring in a less relevant one of the documents.

38. The method of claim 6, wherein the calculated values are weighted based on a ranking defined by relevance of the identified documents, such that an occurrence of a substring in a more relevant one of the identified documents is weighted more than an occurrence of the substring in a less relevant one of the documents.

39. The system of claim 11, wherein the calculated values are weighted based on a ranking defined by relevance of the identified documents, such that an occurrence of a substring in a more relevant one of the identified documents is weighted more than an occurrence of the substring in a less relevant one of the documents.

40. The computer-readable medium of claim 27, wherein the calculated values are weighted based on a ranking defined by relevance of the identified documents, such that an occurrence of a substring in a more relevant one of the identified documents is weighted more than an occurrence of the substring in a less relevant one of the documents.

41. The computer-readable medium of claim 30, wherein the calculated values are weighted based on a ranking defined by relevance of the identified documents, such that an occurrence of a substring in a more relevant one of the identified documents is weighted more than an occurrence of the substring in a less relevant one of the documents.

X. EVIDENCE APPENDIX

XI. RELATED PROCEEDINGS APPENDIX